The Target of Acceptance: Towards a Precise Definition of the Intended Purpose

Dirk Brade
KTH/FOI/FMV
SE-10044 Stockholm, Sweden
dirk.brade@foi.se

Jeroen Voogd
TNO
P.O.Box 96864, 2509 JG
The Hague, The Netherlands
voogd@fel.tno.nl

ABSTRACT: *With the goal to objectively demonstrate the fitness for purpose of a simulation model during its validation, the desire arises to remove vagueness from the terms “fitness” and “purpose”. Several approaches have been presented in the past trying to facilitate and motivate the documentation of "well-defined intended purposes", but still in defense practice a precise purpose specification for a simulation model is rare. Within the methodology developed under the umbrella of WEAG THALES JP11.20 “REVVA”, the concept of the “Target of Acceptance” (ToA) was introduced, enabling clear traceableness between the vague intended purpose and precise, quite objectively assessable acceptability criteria.*
Based on the notion of the Experimental Frame, this paper extends the approach taken in REVVA and proposes a method how to hierarchically derive from the intended purpose a number of sub-objectives of model use, until representative sets of experimental conditions can be defined, under which the simulation model needs to be valid. Analysis of the sensitivity of the contextual objectives to the simulation goal parameters yields an indication for their acceptable imprecision (and thus their acceptable inaccuracy), which can be formulated as testable acceptability criterion.

Keywords: Target of Acceptance, Validity Criteria, Objective Hierarchy, Objective Weakening, Acceptable Residual Uncertainty

1 Introduction

When going to use a simulation model for some intended purpose, it is reasonable to ask the fundamental question, whether this simulation model actually is “fit for this intended purpose”, or whether one has “acquired the right model”. Experience teaches that using an “unfit” simulation model might have unpleasant consequences [1], and, doubtlessly, disastrous consequences must be excluded as far as possible. However, if the above questions are not rephrased more precisely and refined, one will hardly get an objective answer, because neither “fitness” nor “rightness” is directly measurable without further assumptions. As already indicated in [2], each validation effort needs to be based on well-defined criteria to guaranty a minimum desirable degree of objectivity. Today, the assessment of validity of a simulation model is usually based on “Validity Criteria”, but two questions are not yet satisfactory answered:

(1) What do “good” Validity Criteria look like?

(2) How does one get those criteria, and how do they relate to the intended purpose of model use?

This paper addresses these two questions, which are of extra importance in case of planning and implementation of “V&V” by a third party VV&A Agent, as recommended by e.g. [3] and [4]: When a contract is made between the customer party (those who are going to use the simulation model) and the VV&A Agent party (those who are going to assess the simulation model’s validity), good Validity Criteria specify the technical contents of the contract, including both the “what to examine and assess” during V&V and the required rigor of the examination. The concepts proposed here focus on the use for M&S to support analysis and decision making, but it is assumed that they can also be transferred to other areas such as training. Ideally, the proposed activities are performed in the very beginning of a simulation study, before the simulation model is chosen or developed.

The paper roughly outlines a solution approach to one of the many challenges which shall be tackled in the upcoming follow-on project to the WEAG THALES JP11.20 “A Common Verification, Validation, and Accreditation Framework for Simulations” (REVVA). Section 2 introduces the Target of Acceptance (ToA), which is an essential part of the REVVA methodology, and enables the structured documentation of Acceptability Criteria (AC) and their derivation. This section also outlines issues associated with the specification of Validity Criteria (VC), which are as a subset of the set of Acceptability Criteria of special importance for the demonstration of validity. Section 3 then proposes potential solutions to close the gaps identified as most essential, including a very narrow, but precise notion of Validity Criteria based on Observation Sample Points, Behavior Characteristics, and an Experimental Frame Specification (section 3.1), and an approach to dealing with the residual uncertainty that remains even after assessment of a simulation model (section 3.2). Section 4 closes the paper with a summary and conclusions for future work.

2 The Target of Acceptance

Simulation models are going to be used for an intended purpose, and are only considered to be acceptable, if – among other requirements and constraints – they are valid for it. Initially, the intended purpose description must be assumed to be vague, but thus “all-inclusive”. To capture the intended purpose more precisely, to enable the systematic derivation of Acceptability Criteria, and to ensure traceableness between the intended purpose and the Acceptability Criteria, the REVVA methodology features the “Target of Acceptance” (ToA) [5].

The ToA is a hierarchical structure (Directed Acyclic Graph) of objectives and contains as its leaves Acceptability Criteria that the model needs to meet for a particular, well-defined intended use. It answers the question “What exactly needs to be assessed?”, and documents the rationale for the derivation of the acceptability criteria from the intended purpose. On top of the hierarchy stands the vague intended purpose, which is refined into a set of sub-objectives, which again can be decomposed, until Acceptability Criteria related to the M&S product’s correctness and validity can be derived directly from the lowest sub-objectives.

The sub-objectives (child objectives) are supposed to cover their parent objectives without major redundancy, i.e. in such a manner that achievement of the sub-objectives is *necessary and sufficient* for achievement of the higher objective, and finally the intended purpose. Attached to each (sub-) objective is an argument which states why and to which degree the achievement (or the non-achievement) of the child objectives implies the achievement (or the non-achievement) of the parent objective. This argumentation is the glue in the hierarchical structure of the ToA.

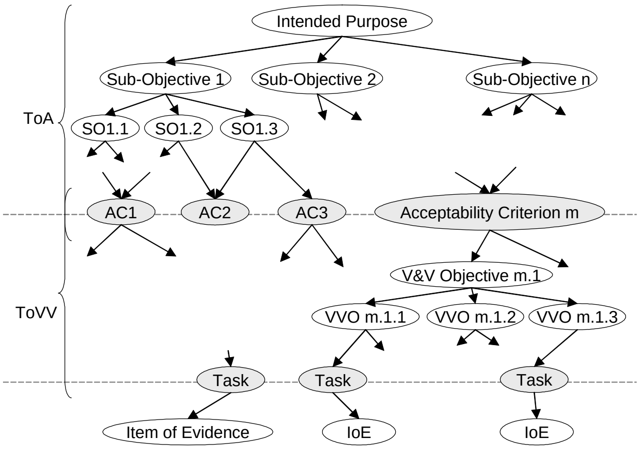

Figure 1: ToA and ToVV

The upper half of Figure 1 (labeled “ToA”) illustrates the concept of objectives decomposition. The REVVA methodology requires that when the V&V effort proceeds, based on the ToA a V&V approach is developed how to demonstrate with the information at hand whether the AC are met or missed [6] (Target of V&V). However, the development of the ToVV is out of scope of this paper, which exclusively concentrates on the improvement of the ToA.

[5] recommends that the assessment of whether an objective stated in a node is met or missed is supported by “measures”. On the higher levels, those measures are expected to be rather abstract, but on the lower levels they should be quite precise. Such measures should not be developed from scratch, but taken from the scientific foundations of the domain from which the objectives originate. For example, battle space simulation offers for the higher levels Measures of Force Effectiveness (MoFE), Measures of Effectiveness (MoE), or Measures of Performance (MoP). Those measures on the highest ToA level are expected to be least directly assessable, while those on the lowest level must be evaluated during the time- and resource-constraint V&V effort.

Although the ToA, which enables clear traceableness between the intended purpose and the AC, promises to be more useful than “naked” AC, it is still lacks the expressiveness to capture all relevant aspects of AC development and maintenance. Not addressed in the REVVA methodology is what Validity Criteria, as subset of the set of AC, should exactly look like, how to deal with objectives that cannot be demonstrated to be achieved (other than to treat them as missed), and how to express the required reliability of the associated V&V activities.

3 Modifications and Extensions to the ToA

In this section the concept of the ToA [7] is revisited, extended, and further formalized. The paper concentrates on refinement and specialization of the intended purpose down to a level at which *behavioral indistinguishableness* between "real" and "simulated behavior" can be determined exclusively based on *behavior characteristics*. The *experimental frame* derived from an elementary sub-objective (i.e., a sub-objective which is not further sub-divided into child sub-objectives) here constitutes the key to the definition of Validity Criteria, which build the foundation for the assessment of the validity of a simulation model. In the following is motivated why good Validity Criteria

1. state *conditions for the accuracy* of behavior characteristics which the simulation model necessarily must satisfy to be valid (section 3.1), and

2. include a statement about the *required reliability* of their associated V&V efforts and the acceptable residual uncertainty of assessment (section 3.2).

As a consequence, the nodes within the ToA are extended. For each node, a limited amount of ranked degrees of weakening shall be identified, and the severity of misjudgment of the achievement of a node's objective estimated and classified in different ranked degrees of negative impact severities (compare [8] or [9]). Those severities then constitute the foundation for the estimation of "how much V&V is enough".

3.1 Valid Behavior

In various explanations or definitions of the term "validation", a relationship to the context of model use is created. While early authors concentrate on the "correctness of the inference about a system derived from the simulation" [10], others later address the "accuracy of a model's behavior within its application domain under consideration of the study objectives" [11], or the "impossibility to distinguish between the system and the model within the experimental frame of interest" [12]. To enable the specification of precise Validity Criteria, and under the assumption that it is hardly possible to prove validity of a nontrivial simulation model, this paper reuses the definitions developed in [5] based on the previous references, and defines:

- Validity (property): The property of a simulation model to have, within a specific experimental frame, a behavior which is indistinguishable from the behavior of the System of Interest.
- Validation (process): The process which is used to construct, under a set of time, cost, skills, and organizational constraints a justified belief about model validity.

Behavior is an abstract concept, and so far has mankind only managed to measure behavior attributes of systems in *observation sample points*. Under the precondition that (dynamic) simulation always is about behavior over time, for validation we exclusively concentrate on simulation model behavior, with the desired objective to demonstrate behavioral indistinguishableness between the simulation model and the system of interest in the relevant observation sample points.

3.1.1 Observation Sample Points

Using a simulation model can be compared to experimentation with a System of Interest (SoI) [12]. One conducts experiments with a SoI to infer some system properties from its behavior observed in samples under deliberately chosen experimental conditions. Observation sample points are defined by what is observed (or better: measured – the *observation attribute*), and the point in time or the time interval the observation takes place (*observation time*). For example, during a live fire test of a torpedo interesting observation sample points might be the torpedo's position in space (observation attribute with three degrees of freedom and a precision of 1m) at a sampling rate of 60Hz (observation times). If simulation shall replace the (real) experimentation, then the simulated system behavior (i.e., the behavior of the simulation model) must be indistinguishable from (i.e., sufficiently similar to) the system's behavior in the observation sample points. We base the demonstration of behavioral indistinguishableness on sets of *behavior characteristics*, which are computed from series of behavior instances observed at these observation sample points.

3.1.2 Behavior Characteristics

Behavior characteristics describe behavior over time in an aggregated form and are computed as functions of behavior instances observed at the observation sample points, recorded during a sufficient number (series) or a sufficiently long experiment (as appropriate). For the purpose of validation, those parameters must be chosen in such a manner that they are useful to *characterize* the SoI's or the simulation model's behavior. An ideal set of behavior characteristics has the property that changes in system behavior, which are relevant for the intended use, are reflected as a change of the value of at least one of the behavior characteristics, and vice versa. A behavior characteristic must be defined in such a manner that it appropriately reflects the degree of variability in system behavior under indistinguishable experimental conditions. For example, the (navigation) behavior of a torpedo might be characterized using a family of trajectories, by its cruise speed, its maximum turning angle, and its operation time.

Behavior characteristics should not be made up for the purpose of validation, unless the application domain does not provide such characteristics. But also if mature behavior characteristics exist in the application domain, care is required, because they rarely truly and completely describe the behavior of the SoI – they are, although helpful, just another model.

3.1.3 Real Experimentation, Simulation, and Experimental Frames

An example of a real experiment is a "live fire test" (observation of a weapon system under deliberately chosen conditions). The idea of experimentation is to deliberately set the conditions under which the SoI behavior is observed or experienced in such a manner that observations made under the experimental conditions yield new insight into SoI behavior, which also can be generalized for other conditions of system operation. The experiment should come with a description of what one would like to achieve with this experiment (e.g., to confirm a hypothesis), and the *experimental conditions* under which it shall take place. For each experiment observation sample points are given as

1. *control parameters* (actively controlled input of the initial values and their change over time),
2. *influence parameters* (passively observed input, but uncontrolled, and their acceptable values over time),
3. additional *instrumentation parameters* (internal, passively observed over time, irrelevant for the experimentation objectives, but relevant for later validation) and
4. *goal parameters* (output, passively observed over time).

These parameters serve to control and perceive in samples the SoI behavior, with this perception being an approximation of the unknown and abstract "real behavior". The *experimental frame* (EF) defines the value domains of the input parameters (control parameters and influence parameters), from which experimental conditions for a set of experiments are chosen. This concept is visualized in Figure 2 [13].

The System of Interest can be hierarchically organized and can consist of numerous subsystems, e.g., humans, machines, or machine components, with own behaviors, which also can be characterized by appropriate behavior characteristics (a fact which later is highly relevant for the planning and implementation of V&V [14]). Experimentation results are gained from the "raw" goal parameter values by post-processing (e.g., statistical evaluation, aggregation, visualization), and often include or exhaustively consist of behavior characteristics.

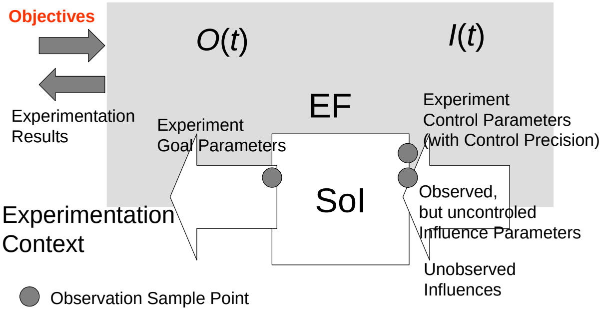

Figure 2: Observation sample points at the System of Interest (SoI) within the Experimental Frame (EF)

Similarly, during (dynamic) simulation, fictive system responses over time are calculated, depending on controlled inputs and the initial internal model state. At given sample observation points control parameter values are injected (for a self-driven model just initially, for a trace-driven or interactive model sequentially), and goal parameter values are recorded over time. The *simulation experimental frame* (SEF) defines the value domains of the control parameters, within which the experimental conditions for a set of simulation experiments are intended to be chosen. This concept is visualized in Figure 3 [13].

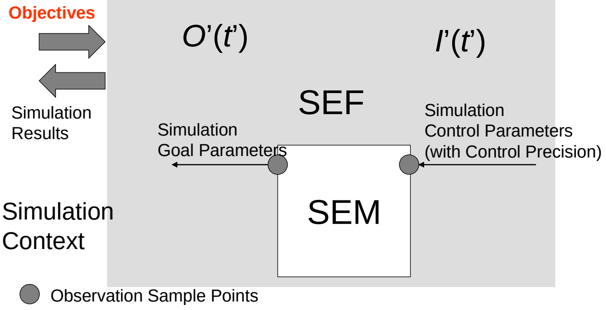

Figure 3: Observation sample points at the Simulation Model (SEM) in its Simulation Experimental Frame (SEF)

When substituting real experimentation by computer simulation, the intended purpose does not change. The observation sample points of the real experiment are likely to be different from the observation sample points in the simulation experiment, but for the remainder of this paper is assumed that bidirectional functions can be defined, which project the observation sample points of the real experiment to those of the simulation experiment and vice versa. We also assume that it is possible to functionally project the simulation goal parameters to the real experiment goal parameters.

The simulation results (post-processed simulation goal parameters) must not deviate from the experimentation results (as they are or could be recorded at the SoI) in a harmful manner. This is likely, if a series of observation trajectories gained at the observation sample points during multiple instantiations of experiments within a given EF cannot be distinguished from the series of observation trajectories of the associated observation sample points in the associated comparable real experiments within their given range of precision. If one characterizes the series of observation trajectories using behavior characteristics, the acceptable imprecision of the associated "real" behavior characteristics serves as the threshold of indistinguishableness.

3.1.4 Acceptable Imprecision of the Behavior Characteristics

The main challenge for modeling and simulation is to approximate the experiment as it could take place in reality as defined by the experts from the application domain as accurately as necessary. For a real experiment, the required precision for the input parameters (control and influence parameters) depends on the sensitivity of the goal parameters to the individual inputs. For decision making or analysis, the variability of the goal parameter values must be sufficiently low, placing implicit precision requirements on the control parameters and the simulation model itself. Here it is assumed that the experts within the application domain have methods and tools available to estimate an upper bound of this required precision. The required precision of the goal parameters depends on the sensitivity of the Simulation Results to the individual goal parameters. For their determination, again knowledge from the application domain is required.

The observation attribute dimension of a VC states that the SEM behavior characteristic c' does not deviate from the SoI behavior characteristic c more than the given threshold d, given the difference function or distance measure δ.

Then is can be expressed as 4-tuple

$$< c, c', \delta, d >$$

or by an expression such as

$$d_{lower} \leq \delta(c, c') \leq d_{upper}$$

The work to identify relevant behavior characteristics, an appropriate distance measure, and the threshold constitutes a key piece of a validation effort and requires rather expertise in the application domain than M&S knowledge.

3.2 Required V&V Reliability

For various reasons it does not seem to be possible to doubtlessly prove the validity of a simulation model, which opens the possibility that the perception of validity (or invalidity) is wrong. Two classes of uncertainty within any set of AC and their substantiating material can be clearly categorized when using the concept of the ToA:

1. Uncertainty within the nodes: The uncertainty associated with the assessment of the individual Validity Criteria: To substantiate a VC, Items of Evidence are assembled. Based on the reliability of the Items of Evidence (which is measured as probative force) and the necessity and sufficiency of the items of evidence to back-up the VC (which is measured as convincing force), there comes some uncertainty with the assessment result of each VC.
2. Uncertainty among the nodes: The ToA is a hierarchical structure, which decomposes the intended purpose description into SO's and finally into experimental conditions. Here the degree of sufficiency and necessity of the set of experimental conditions (bottom of the hierarchy) with respect to coverage of the intended purpose (top of the hierarchy) is uncertain.

Accepting that this uncertainty can not be eliminated, an estimate of the residual uncertainty which helps to consider it appropriately for the acceptance/rejection decision becomes desirable.

3.2.1 Erroneous Assessment

At any point in time there remains some uncertainty that a simulation model, which is perceived as correct and valid for its intended purpose, factually is not (and vice versa). Table 1 abstractly depicts this problem assuming a binary "valid – invalid" classification. If a simulation model is perceived as invalid, there is some uncertainty whether this perception might be a "false alarm". If a Simulation Model is perceived as valid, there remains some uncertainty that the perception deviates from the fact.

Table 1 does not only apply to the overall "fitness for purpose", but also for all sub-nodes in the ToA (objective achieved/not achieved). Having the intended purpose, its sub-objectives, and experimental conditions defined in the ToA, for all nodes in the hierarchy uncertainty estimates associated with the statements made can be given.

From the decision maker's perspective, among the most popular VV&A challenges is to "balance the risk of assuming against the cost of knowing" [15]. This expression can be reformulated (less "snappy") and refined to: "The severity of a negative impact caused by wrong assessment of a simulation product must be indirectly proportional to the uncertainty associated with the validity assessment". In general, as a consequence of non-provability of validity, any acceptance decision concerning a simulation model is made on uncertain information. For responsible decision making, an estimate of the upper bound of this uncertainty is required.

Table 1: Factual validity vs. perceived validity

Fact Perception (Action)	Factually valid	Factually invalid
Perceived as valid (and accepted)	Ok	Type II Error
Perceived as invalid (and rejected)	Type I Error "False Alarm"	Ok

3.2.2 Weakening of Objectives

To consider validity as a binary property is theoretically sound, but not satisfactory in practice. For this purpose the concept of *relaxation* or *weakening* of an objective is introduced, which is related to the "must/should/can approach" of Software Engineering. While the simulation model is not valid for the original intended use, it might very well be valid for a weaker intended purpose. Here we propose to explicitly identify discrete degrees of *objective weakening*. It looks promising to proceed stepwise by defining thresholds for the *complete miss* of a node objective, a *major weakening* of the node objective, and its *minor weakening*. The choice of two interim weakening levels made here is arbitrary, and it is up to the decision maker, to choose more or less steps. The weakening of the required precision of the observation attribute within a VC is a relaxation of the tolerance border *d*:

$$d_{weaker\ \ lower} \leq d_{stronger,\ lower} \leq \delta(c, c') \leq$$

$$\leq d_{stronger,\ upper} \leq d_{weaker,\ upper}$$

Degrees of weakening must be defined consistently in the ToA all over its nodes from the Validity Criteria up to the intended purpose. Here it is recommended to define the weakening of a higher node's objective based on a (logical) expression of the weakening of its child nodes objectives bottom up (e.g., Table 2).

The identification of degrees of objectives weakening does not only play an important role for the stabilization of the objectives hierarchy, but it is likely that at least some of them will serve as fall-back solution, when it turns out that under time and budget constraints it is not possible to make sufficiently sure that the objective indeed is achieved. Then the V&V effort can be concentrated on demonstrating that "at least" a particular weakening level is met.

Table 2: Parent objective SO1 weakened by degree of achievement of child objectives SO1.1 and SO1.2

SO1.1	SO1.2	SO1
Full achievement	Full achievement	Full achievement
Full achievement	Minor Weakening	Full achievement
Minor Weakening	Full achievement	Full achievement
Full achievement	Major Weakening	Minor Weakening
Minor Weakening	Minor Weakening	Minor Weakening
Major Weakening	Full achievement	Minor Weakening
Full achievement	Complete Miss	Major Weakening
Minor Weakening	Major Weakening	Major Weakening
Minor Weakening	Complete Miss	Major Weakening
Major Weakening	Minor Weakening	Major Weakening
Complete Miss	Full achievement	Major Weakening
Complete Miss	Minor Weakening	Major Weakening
Major Weakening	Major Weakening	Complete Miss
Major Weakening	Complete Miss	Complete Miss
Complete Miss	Major Weakening	Complete Miss
Complete Miss	Complete Miss	Complete Miss

3.2.3 Impact and Acceptable Residual Uncertainty

The risk associated with an erroneous V&V assessment is considered to be the driver for V&V rigor. A threshold for the *acceptable residual uncertainty* (ARU) for the assessment of each objective within the ToA shall be set, based on the severity of the negative impacts stated. As a rule of thumb, the higher the negative impact of wrongly assessing whether the simulation model is suitable to achieve an objective in the hierarchy, the lower the acceptable residual uncertainty. Consequences for both overestimation and underestimation must be considered. It is up to the acceptance decision maker or an organization's policy (e.g., risk averse, risk neutral, risk friendly, as used in Operations Research [16]) to define the acceptable threshold on the given uncertainty scale for a particular degree of severity.

To rank lingual severity statements, impact domains have been identified by [8] and [9], which serve here as factors for the estimation of the worst case impact. They quantify the potential consequences of the use of erroneous M&S product in four distinct classes for all identified impact domains, ranging from "negligible", to "marginal", "critical", and "catastrophic" and give examples.

Table 3 shows an example of estimated severities of misjudgments for sample sub-objective 1 (SO1) "Analyze intrusion detection capability" from an M&S user perspective. The upper right of the matrix shows the severity of the impact of applying the simulation model for SO1 while overestimating the degree of objective achievement. It states, e.g., that it would be “catastrophic”, if the results that the simulation model produces are far from reality, although V&V attests that the simulation model is valid. (The rating “catastrophic” is chosen, because during implementation of the analysis results based on this simulation study the navy’s capability to detect submarine intruders will be reduced for a whole acquisition cycle.) The lower left matrix shows the severity of the impact of not fully leveraging the potential of the simulation model for SO1 due to its underestimation. It states, e.g., that the severity of the consequences are “marginal”, if the simulation model is perceived as completely unsuitable to examine intrusion detection questions, although it factually is. (This severity rating is made based on the assumptions that the development of the simulation model is relatively cheap and there is no time-critical thread imaginable in near future.)

Table 3: Estimated severities of misjudgments for SO1

<table>
<tr><td></td><td colspan="5">Fact</td><td></td></tr>
<tr><td rowspan="5">Perception</td><td>SO1</td><td>Fully achieved</td><td>Minor weakening</td><td>Major weakening</td><td>Complete miss</td><td rowspan="5">Overestimation (type II error)</td></tr>
<tr><td>Fully achieved</td><td>Ok</td><td>Marginal</td><td>Critical</td><td>Catastrophic</td></tr>
<tr><td>Minor weakening</td><td>Negligible</td><td>Ok</td><td>Critical</td><td>Catastrophic</td></tr>
<tr><td>Major weakening</td><td>Negligible</td><td>Negligible</td><td>Ok</td><td>Critical</td></tr>
<tr><td>Completely missed</td><td>Marginal</td><td>Marginal</td><td>Negligible</td><td>Ok</td></tr>
<tr><td></td><td colspan="5">Underestimation (type I error)</td><td></td></tr>
</table>

3.2.4 Measures of Uncertainty

Three extremes for the qualification of the truth of a statement are distinguished: *Prove* (no doubt about truth of statement), *disprove* (no doubt about falseness of statement), and *complete ignorance* (don’t know whether the statement is true or false). With incomplete information available, the first two extremes are rare, and the third is of special importance in case of information absence or contradicting, equally strong items of information. Quantified uncertainty serves as a measure of distance of statement certainty from the state of complete ignorance. With the abilities to quantitatively express uncertainty, to estimate the residual uncertainty which is left after V&V, and to define a threshold for the acceptable residual uncertainty, the question “how much V&V is enough” can be answered. For the proposed approach, thresholds for the ARU shall be derived directly from the severity statements.

An appropriate measure of uncertainty must be easily understandable and come with a consistent body of rules and operations for its processing. The semantics should be intuitively understood and appropriately reflected in the propagation rules. Measures for uncertainty are numerous (e.g., belief function [17], possibility theory [18], probability theory), but often their appropriateness in the given context is questionable. Such which have been reviewed for M&S VV&A by the author include linguistic ranked variables [19], possibility measures [13], and probability measures [20]. This paper does not favor a particular measure for V&V uncertainty, but stresses the fact that a measure appropriate for the individual ToA development and contract specification needs to be chosen and understood.

4 Summary and Conclusions

This paper addressed the problem of how to traceably develop and to unmistakably define strict Validity Criteria. For this purpose the concept of the Target of Acceptance (ToA) introduced by the WEAG THALES JP11.20 “A Common VV&A Framework for Simulations” (REVVA) was revised, which encourages the hierarchical decomposition of the intended purpose of model use into sub-objectives and sub-sub-objectives, until experimental frames to achieve the lowest objectives can be derived. Based on a strict definition and separation of simulation model validity and validation, two-dimensional Validity Criteria were introduced, consisting of a precision condition for an attribute and an expression of the acceptable residual uncertainty (ARU) for the associated assessment activity. To overcome the complications associated with a binary validity/invalidity assessment, now objectives within the ToA may be stepwise weakened, which facilitates the demonstration of the achievement of a weaker objective, if the simulation model is heading for rejection with respect to the originally intended purpose.

The effort associated with this process of requirements elicitation is considered to be high, but should pay back for simulation models of higher complexity. The presented approach will be refined an developed in more detail during the follow-on program of REVVA, which is scheduled to start during Spring 05.

Acknowledgements

The presented approach has been developed within the context of a research project sponsored by the Swedish Defense Research Agency (FOI) and the Defense Material Administration (FMV). All reports produced by the WEAG THALES JP11.20 consortium can be freely accessed and downloaded under www.jp1120-revva.com. Colin Brain (SEValidation), Martijn van Emmerik (TNO), René Jacquart (ONERA), Fredrik Jonsson

(FMV), and Håkan Lagerström (FMV) were actively involved in the evolution and stabilization of the concepts presented in the paper.

References

[1] Frankfurter Allgemeine Zeitung. 1995. Amerikanische Behörde warnt vor Fehlern in der Naviagationselektronik, 12. August.

[2] Balci, O. 2000. Well-Defined Intended Uses: An Explicit Requirement for Accreditation of Modeling and Simulation Applications, Proceedings of the 2000 Winter Simulation Conference, pp. 849-854, Society for Modeling and Computer Simulation International.

[3] Arthur, J.D. and R.E. Nance. 2000. V&V Without Independence: A Recipe for Failure. In Proceedings of the 2000 Winter Simulation Conference, Society for Modeling and Computer-based simulation International.

[4] Rosenberg, L.H. 2001. Technical Report: Verification and Validation Implementation at NASA. NASA, Goddard Space Flight Center.

[5] METHGU2. 2004. THALES JP11.20 Report JP1120-WE5100-D5101-METHGU2-V1.2.

[6] PROSPEC. 2004. THALES JP11.20 Report JP1120-WE5200-D5201-PROSPEC-V1.3

[7] TOAGUID. 2004. THALES JP11.20 Report JP1120-WE2200-D2201-TOAGUID-V1.1.

[8] Muessing, P., D. R. Laack, and J. J. Wrobleski, Jr. 1997. Optimizing the Selection of VV&A Activities: A Risk/Benefit Approach, Proceedings of the 1997 Summer Computer Simulation Conference, Society for Computer Simulation International.

[9] Mugridge, C. 1999. Verification, Validation and Accreditation of Models and Simulations Used for Test and Evaluation – a Risk/Benefit Based Approach. Internal Report, Technical Development Group, Defense Evaluation and Research Agency UK.

[10] Shannon, R.E. 1975. Systems Simulation and the Art of Science. Prentice Hall, Eaglewood Cliffs, N.J.

[11] Balci, O. 1998. Verification, Validation and Testing. In: The Handbook of Simulation edited by J. Banks, Wiley, New York.

[12] Zeigler, B.P., H. Praehofer, and T.G. Kim. 2000. Theory of Modeling and Simulation. Second Edition. Academic Press. ISBN: 0-12-778455-1

[13] Brade, D. 2004. Quantitative Uncertainty Metrics in M&S VV&A. Foundations ´04: A Workshop for VV&A in the 21st Century, Tempe, AZ.

[14] Flight Mechanics Panel Working Group WG-12 on Validation of Missile Simulation. 1985. Final Report, AGARD Advisory Report No. 206.

[15] Brain, C. 2001. Achieving the Right Level of Credibility, Presentation at the JASA Workshop: Assuring M&S Credibility for Defense Acquisition and T&E, Reno.

[16] Hiller, F.S. and G.J. Liebermann. 1997. Operations Research – Einführung. 5. Auflage, Oldenbourg-Verlag München Wien, ISBN: 3-486-23987-2.

[17] Shafer G. 1976. A Mathematical Theory of Evidence, Princeton University Press.

[18] Dubois, D. and H. Prade. 1988. Possibility Theory: An Approach to Computerized Processing of Uncertainty. Plenum Press, New York and London.

[19] Brade, D. and A. Köster. 2001. Risk-based Validation & Verification Levels Definition. Proceedings of the European Simulation Interoperability Workshop, London, Simulation Interoperability Standardization Organization.

[20] Brade, D. and M. Lemaître. 2004. Assessment Criteria and V&V Levels in JP11.20. Spring Simulation Interoperability Workshop, Simulation Interoperability Standardization Organization (SISO), Arlington, VA.

Authors Background

Dr. Dirk Brade is employed as guest researcher at the Royal Institute of Technology (KTH) in Stockholm on a position sponsored by the Swedish Defense Research Agency (FOI) and the Swedish Defense Materiel Administration (FMV). He received a Doctor Degree in Natural Sciences from the University of the German Federal Armed Forces Munich (UniBwM) in 2004, and a Diploma Degree in Computer Science from the Technical University Darmstadt in 1998. He participates in several international VV&A working groups on behalf of and in close cooperation with FOI and FMV. His research interest focuses on VV&A, decision making, component based simulation, the HLA, and formal modeling languages.

Dr. Jeroen Voogd is a member of the scientific staff in the Defence, Security and Safety Division at TNO. He holds a Ph.D. (1998) in Computational Physics from the University of Amsterdam in the field of modeling and simulating biophysical systems on parallel and distributed computing platforms. Currently he is involved in projects on simulating group behavior, operational analysis studies of army operations and connecting C4I infrastructure to simulators. A recurring theme in his work of the last years is quality. This involves issues round fidelity and VV&A of simulator assets, as well as quality assurance within TNO in the form of e.g. audits.